Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

A Hybrid CNN-Based Attention Model for Weed Recognition Under Variable Field Conditions with Edge Device Simulations

Dharmendra Kumar[1], Sunil Dhankhar[2], Saroj Hiranwal[3], Pushpendra Singh Sisodia[4]

[1]Department of Computer Science and Engineering, Rajasthan Institute of Engineering and Technology, Bhankrota, Ajmer Road, Jaipur, Rajasthan, India.
[1]Department of Computer Science and Engineering, Anand International College of Engineering, Near Kanota, Agra Road, Jaipur, Rajasthan, India.
[2]Department of Computer Science and Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur, Rajasthan, India.
[3]Course Director-IT Programs, Australian Institute of Advanced Technologies, Adelaide, SA 5000, Australia.
[4]Department of Artificial Intelligence, Topia Life Sciences, Ahmedabad, Gujarat, India.

E-mail: [1]dharm.1982@gmail.com, [2]sunil@skit.ac.in, [3]ersarojhiranwal@gmail.com, [4]pushpendra.muj@gmail.com

Orcid ID: [1]0009-0005-8130-590X, [2]0000-0003-4600-267X, [3]0000-0002-3680-3476, [4]0000-0001-5140-5743

Abstract

Biotic stress significantly affects agricultural productivity, and weed infestation remains a major factor limiting crop yields and resource-use efficiency. Traditional weed management tools and practices are less effective and labor-intensive at scale. Therefore, automated and intelligent weed identification and classification technologies are required to manage weed infestation. Recent deep learning studies have shown promising results for weed classification. Still, many deep learning architectures have large parameter counts and require high computational power, limiting their deployment in real-world agricultural environments. In this study, a novel lightweight deep learning model, Lightweight Hybrid Channel-Spatial Attention (L-HCSA), is presented for crop-weed image classification in a real-world agricultural environment. L-HCSA is built on a pre-trained MobileNetV2 backbone, enhanced with a hybrid channel-spatial attention module to handle background noise and enhance the discriminative power of weed features. The CWD30 database is used to benchmark the L-HCSA model. An ablation study was performed on five configurations, confirming that the HCSA attention block is the primary contributor to performance. The HCSA attention block adds 0.41 million parameters to the MobileNetV2 baseline (2.63 million parameters) and improves accuracy by 2.18%. The L-HCSA is compared against nine baseline models, including MobileNetV2, MobileNetV3-L, DenseNet-121, ResNet-50, ResNet-101, EfficientNetV2-M, EfficientNet-Lite, TinyViT, and EdgeNeXt. The L-HCSA model has only 3.04 million trainable parameters (12.1 MB) and achieves 77.34% accuracy with a ROC-AUC of 0.9955. A pairwise Wilcoxon signed-rank test with Bonferroni correction and Cohen's d effect size was performed for statistical evaluation. The L-HCSA model is evaluated for real-time performance, meeting the >=10 FPS threshold in an embedded hardware simulation while maintaining competitive accuracy, achieving 22.0 FPS on the Google Coral Edge TPU.

DOI: https://doi.org/10.36548/jiip.2026.3.010
Received: 09.05.2026, received in revised form: 15.06.2026, accepted: 04.07.2026, published: 14.07.2026

Keywords: Crop Weed Classification, Deep Learning, Precision Agriculture, Attention Mechanism, and Computer Vision.

1. Introduction

Agriculture is a major driver of the Indian economy and plays an important role in sustaining human activity for generations. The agricultural sector has evolved over the last few decades through technological advancements that have strengthened the Indian economy. This technological transformation makes India a leading producer of cereals, pulses, fruits, and vegetables worldwide [1]. The integration of digital technologies, such as Remote Sensing, the Global Positioning System (GPS), and data-driven decision support systems, has significantly improved agricultural productivity [2]. Also, long-term assessments of India's agricultural transition show structural improvements, including mechanization, improved irrigation, and high-yielding crop varieties [1]. Despite technological and structural advancements, Indian agriculture faces major productivity constraints due to biotic stresses. Biotic stress, such as weed infestation, is the most serious limitation to crop productivity, especially in major cereal, oilseed, and pulse-growing regions. Weeds aggressively compete for light, water, nutrients, and space, reducing plant vigor, altering canopy structure, and lowering harvestable yields [3]. Research indicates that weed competition alone impacts the Indian economy to lose 11 billion dollars annually. Weeds can reduce crop yield by 30% to 76%, depending on crop type, weed flora, and management practices [4]. Nutrient depletion in soil due to weed competition can directly impact staple crop production, such as wheat [5]. Despite these challenges, weed management strategies such as precision agriculture tools, improved weed surveillance, and data-driven approaches are important for agricultural growth. These tools and technological advancements are expected to reduce biotic stress and support long-term agricultural resilience [6]. Recent studies show that rapid progress in agricultural automation is increasing the focus on developing intelligent systems capable of distinguishing plant species, especially weeds, which significantly affect crop yield and productivity. Digital technological advancements, such as deep learning and machine learning, have emerged as powerful tools for precisely identifying weeds. Automated weed classification using deep learning models improved efficiency by learning complex patterns that traditional algorithms struggle to capture. Deep learning models, especially VGGNet, can accurately classify diverse weed species from high-resolution field images [7]. Another study introduces a weed detection and mapping system using a YOLOv8-based architecture that integrates with an actuator trajectory algorithm to precisely optimize weeding path planning at lower computational cost [8]. Such models are valuable for monitoring large agricultural fields where manual inspection is inefficient. However, the model's reliability depends on carefully curated large datasets, appropriate model architecture selection, and effective training strategies. Incorporating diverse data augmentation strategies, such as variations in lighting, viewpoint, and background clarity, ensures the model’s robustness under real farming conditions. A recent study indicates that such preprocessing, feature engineering, and augmentation strategies are essential for improving the model’s generalization and deployment capabilities for weed classification in varied agricultural environments [9].

In this study, we introduce a novel lightweight deep learning-based architecture termed Lightweight Hybrid Channel-Spatial Attention (L-HCSA). The L-HCSA model is developed specifically for agricultural edge deployment. L-HCSA applies a hybrid attention block that uses channel attention and spatial attention in a sequential pipeline to the MobileNetV2 backbone. The performance of the L-HCSA model is compared with previously used deep learning models such as ResNet-101 [10], MobileNetV2 [11], MobileNetV3-L [12], EfficientNetv2-M [13], DenseNet-121 [14], ResNet-50[15], EfficientNet-Lite [16], TinyViT [17], and EdgeNeXt [18]. Several important observations are highlighted in this study through a comprehensive ablation study, statistical testing, and edge deployment simulation across four hardware platforms, providing the empirical grounding needed to assess both the model's contributions and its practical deployment readiness.

The paper is organized as follows: Section 2 (literature review) presents related prior work, deep learning technologies used for weed classification, and research gaps. Section 3 provides a detailed description of the dataset used in the study. Section 4 explains the architecture and workflow of the proposed model, L-HCSA, as well as its training procedures. In Section 5, we report our experimental results. Finally, Section 6 concludes the paper and outlines future research.

2. Related Work

Initial studies primarily relied on CNNs for weed identification, demonstrating that standard architectures could reliably differentiate crops from weeds across several cropping systems. Deep CNNs were further employed for detection tasks in UAV imagery, allowing large-scale scouting and mapping of weed infestations. More recent studies have explored optimized, lightweight architectures tailored for high-speed field inference, including FPGA-accelerated models and efficient YOLO variants for edge deployment [19]. Deep learning has become the foundation of modern crop-weed recognition systems, driven by the availability of large annotated datasets. Several publicly available datasets, such as CWD30 [20] are used for performance benchmarking and model evaluation for real-world agricultural scenarios

These datasets enable robust model development under practical constraints by capturing diverse field conditions, the distributions of multiple weed species, and varying illumination levels.

However, most datasets suffer from class imbalance and environmental noise. Data augmentation techniques, such as image translation, hyperspectral augmentation, and multiple-fold sampling, help improve the model's classification performance. For scenarios where labelled images are scarce, unsupervised and semi-automatic labelling frameworks have been investigated to expand training datasets with minimal manual intervention.

Recent studies indicate that transfer learning has played a crucial role in improving classification accuracy with limited agricultural data. Studies evaluating deep transfer learning for identifying weeds in cotton, soybean, strawberry, and pea fields showed substantial improvements over training-from-scratch approaches [21]. Similarly, transformer augmented and multi-scale feature models have demonstrated improved generalization in complex environments. DenseNet variants, feature-enhancement modules, and multi-class discriminative architectures were applied to sorghum, corn, greenhouse crops, and vegetable systems. Beyond classification, several studies have integrated deep learning into robotic and real-time agricultural systems. Under-canopy robotic weed control and Unmanned Aerial System (UAS)- based object detection pipelines demonstrated practical paths toward automated precision spraying and weed management. Additionally, new datasets and benchmarking initiatives, such as iNatAg and the Weed-Crop Dataset for robotic systems, further expanded the scope of multi-species weed recognition across diverse environments [22].

Overall, existing research highlights major progress in crop-weed detection, yet persistent challenges remain. Many models struggle in visually ambiguous scenarios such as narrow leaf grass vs crop seedlings, multi-plant occlusions, and highly imbalanced weed distributions. Computational overhead remains a barrier for real-time deployment in low-power field devices. These gaps motivate the development of more efficient, attention-enhanced deep networks tailored specifically for weed classification under real-world field variability.

2.1 Gaps Identified in Current Literature

1. First, there aren't many ablation experiments in agricultural deep learning that focus on the impact of specific architectural elements, especially attention mechanisms.
2. Second, evaluating true architecture trade-offs is challenging because comparisons between lightweight models like MobileNetV2 and compact vision transformers like TinyViT or EdgeNeXt on the same benchmark dataset are uncommon.
3. Third, statistical evaluation with repeated experimental runs is largely missing from agricultural deep learning studies.
4. Fourth, despite repeated assertions of edge suitability, proof of such suitability (simulated hardware benchmarks on Jetson Nano, FPGA, or TPU devices) is rarely presented.
5. Fifth, in the context of limited edge hardware simulation, the majority of studies report accuracy without considering latency, throughput, or model size.

Collectively, these gaps suggest that current approaches do not fully address the trade-offs among parameters, accuracy, latency, throughput, size, performance, and practical deployment constraints for precision agriculture edge devices.

3. Dataset Description

To comprehensively evaluate the proposed L-HCSA model for weed classification, the benchmark dataset CWD30 is used to assess its generalization.

3.1 CWD30 Dataset

Serving as the primary benchmark for model generalization and state-of-the-art comparisons, CWD30 is a comprehensive, hierarchical dataset introduced in 2023 and refined in a 2025 publication [26]. The CWD30 dataset comprises 30 fine-grained classes and 219,770 high-resolution RGB images. These 30 classes are divided into 10 crop species and 20 ecologically diverse weed species. Crop species include Corn, Soybean, Lettuce, Radish, Wheat, Rice, Tomato, Potato, Cotton, Sorghum, whereas, ecologically diverse weed species include Bluegrass, Lamb's Quarters, Sedges, Spiny Sow Thistle, Pigweed, and various Broadleaf and Grassy invasive species. Images are captured from Canon EOS and DJI Phantom UAVs from fields across Asia, Europe, and North America. The CWD30 dataset spans multiple seasons, growth stages, viewing angles, lighting variations, and real-world interference like occlusion and residues. The dataset is split 60% crops and 40% weeds to simulate field conditions. For model evaluation, the dataset is split into 70%, 15%, and 15% for training, testing, and validation, respectively.

4. Methodology

To provide high recognition performance and low computational cost for resource-constrained agricultural applications, a Lightweight Hybrid Channel-Spatial Attention (L-HCSA) model is proposed in this study. The architecture of L-HCSA integrates a pre-trained MobileNetV2 backbone with a hybrid channel-spatial attention mechanism and a compact classification head. The complete architecture of L-HCSA is illustrated in Figure 1.

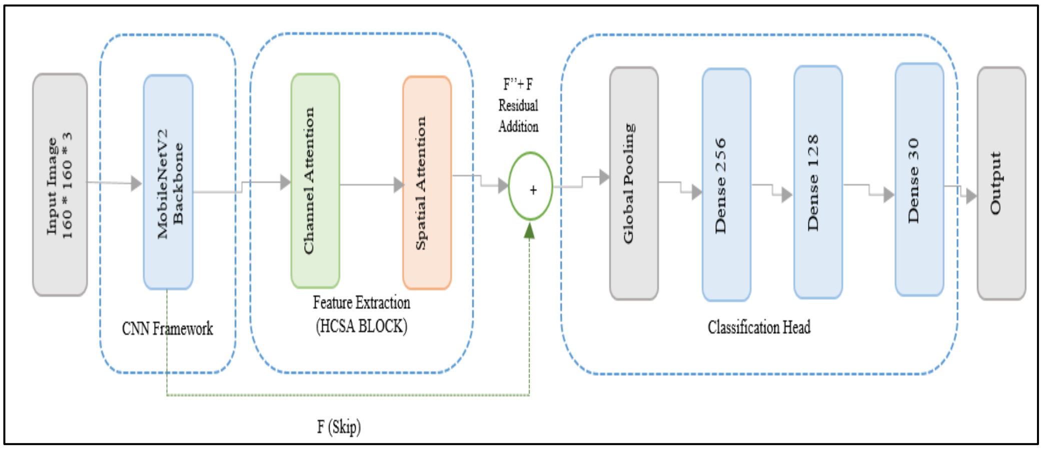

Figure 1. Architecture of the proposed L-HCSA model, comprise CNN framework, feature extraction and classification pipeline

4.1 MobileNetV2 Backbone

MobileNetV2 serves as the feature-extraction backbone for the proposed L-HCSA model. It uses inverted residual blocks and depth-wise separable convolutions, resulting in 8-9 times fewer multiplication and addition operations than typical convolution layers. MobileNetV2 uses an input picture $x \in R^{160 \times 160 \times 3}$ to generate a feature map $F \in R^{5 \times 5 \times 320}$. Where x is the input image, F is the feature map, and R is the tensor. As illustrated in Figure 2, the first 60% of backbone layers are frozen during training to preserve generalized ImageNet visual priors, while the remaining 40% are fine-tuned on the agricultural dataset. The backbone generates 2.63 million parameters at 0.01 GFLOPs.

4.2 Hybrid Channel-Spatial Attention (HCSA) Block

The HCSA block uses channel attention first, then spatial attention in a sequential pipeline.

4.2.1 Channel Attention

Channel attention selects the most informative feature channels and reweights them before spatial analysis. Two-channel descriptors are generated from the feature map $F \in R^{H \times W \times C}$ using global average and global max pooling. Where H stands for Height, W stands for Width, and C stands for Channels. A shared two-layer MLP then processes each descriptor.

$$M_c(F) = \sigma\left(W_2 \cdot ReLU\left(W_1 \cdot AvgPool(F)\right) + W_2 \cdot ReLU\left(W_1 \cdot MaxPool(F)\right)\right) \quad (1)$$

Where $W_1 \in R^{C/r \times C}$, $W_2 \in R^{C \times C/r}$ are shared MLP weights, reduction ratio $r = 8$, σ represents sigmoid activation, and $M_c(F)$ is the channel attention map function in eq. (1). The channel-attended output formula is: $F' = F \odot M_c(F)$. Where F' is the channel–attended feature map, and $\odot$ is element-wise multiplication.

4.2.2 Spatial Attention

After concatenating channel-wise average and max pooling, a 7x7 convolution is applied to choose the most informative spatial regions inside F'. Where $M_s(F')$ is the spatial attention map in eq. (2).

$$M_c(F') = \sigma\left(f^{7\times7}\left([AvgPool_c(F'); MaxPool_c(F')]\right)\right) \quad (2)$$

The spatially attended output is: $F'' = F' \odot M_s(F')$. The HCSA block adds about 0.41 million parameters to the backbone.

4.3 Residual Addition

A residual addition is introduced between the MobileNetV2 backbone and the Hybrid Channel-Spatial Attention (HCSA) block to ensure the HCSA mechanism refines rather than replaces the original feature representation. The residual addition is denoted by $(F'' + F)$. Here, F denotes the original feature map produced by the MobileNetV2 backbone, and F'' is the feature map produced when F is passed through the HCSA block. The residual addition $(F'' + F)$ combines the refined feature with the original skip-connected feature map, preserving information from both the MobileNetV2 backbone and the HCSA block.

4.4 Classification Head

Global Average Pooling reduces F'' to a 1D descriptor. Two fully connected layers (FC-256 to FC-128) follow, with Batch Normalizations, ReLU activation, and Dropout (p = 0.40 and 0.30, respectively). The final prediction employs Softmax over 30 classes.

4.5 Loss Function and Training

The model is trained end-to-end using categorical cross entropy and L_2 weight decay ($\lambda = 10^{-4}$):

$$L = -\frac{1}{N}\sum_i \sum_c y_{i,c} \log \widehat{p_{i,c}} + \lambda \|W\|_2^2 \quad (3)$$

The loss function has two components as described in Eq. (3)-

The first part is categorical cross-entropy, given by $-\frac{1}{N}\sum_i \sum_c y_{i,c} \log \widehat{p_{i,c}}$. Where N is the total number of training samples, i is the index of each sample, c is the Index of each class, $y_{i,c}$ is the ground truth and $\log \widehat{p_{i,c}}$ is the predicted probability that sample i belongs to class c. The second part is L_2 weight decay, which is represented by $\lambda \|W\|_2^2$. Where λ is the regularization strength, W is all the learnable weights in the network, and $\|W\|_2^2$ is the sum of squares of all weights.

The Adam optimizer is employed with an initial learning rate of 1×10^{-3}, which is reduced by a factor of 0.5 upon reaching the validation loss plateau (patience = 4, minimum $lr = 1 \times 10^{-7}$). Early halting (patience = 8 epochs) based on peak validation accuracy reduces overfitting. All experiments run on a Kaggle NVIDIA Tesla P100 GPU (16 GB HBM2) with an image size of 160×160 and batch size of 16. Figure 2 depicts the workflow of the proposed Lightweight Hybrid Channel-Spatial Attention (L-HCSA) model.

4.6 Wavelet-Based Ablation Variants

The primary L-HCSA model integrates a MobileNetV2 backbone with a Hybrid Channel-Spatial Attention (HCSA) block to refine extracted features, as described in Section 4.1 and 4.2. To further investigate whether frequency-domain information could enhance or replace this HCSA mechanism, we introduce three additional ablation variants that incorporate wavelet-based decomposition. The first variant applies a MobileNetV2 backbone with wavelet decomposition directly to its features, without any subsequent attention refinement, thereby isolating the standalone contribution of frequency-domain decomposition. In this variant, the backbone feature map is decomposed using the Discrete Wavelet Transform (DWT) into approximation (low-frequency) and detail (high-frequency) sub-bands [23]. The second variant combines the MobileNetV2 backbone with Haar Wavelet and an attention mechanism. The Haar wavelet is computationally efficient at capturing abrupt intensity transitions, such as object edges, and is followed by the attention modules [24]. The third variant combines the MobileNetV2 backbone with the Daubechies-4 (DB4) wavelet and an attention mechanism. The DB4 wavelet employs smoother, overlapping basis functions that represent more gradual textural variations at a slightly higher computational cost, followed by an attention mechanism [25]. All three wavelet-based variants are evaluated solely in Section 5.1 (Ablation Study) to isolate the individual and combined effects of frequency-domain decomposition and attention on classification performance, relative to the L-HCSA and the baseline MobileNetV2 backbone.

4.7 Statistical Evolution Protocol

To ensure statistical reliability, all models are trained and evaluated over three separate runs using random seeds {42, 123, 777}. All measures are presented with their mean, standard deviation, and 95% confidence interval ($CI = 1.96 \times SEM$). To determine the statistical significance of differences in accuracy, pairwise Wilcoxon signed-rank tests (two-sided) with a Bonferroni correction are used across all model pairs. The protocol also included Cohen's d to quantify the practical magnitude of differences in accuracy.

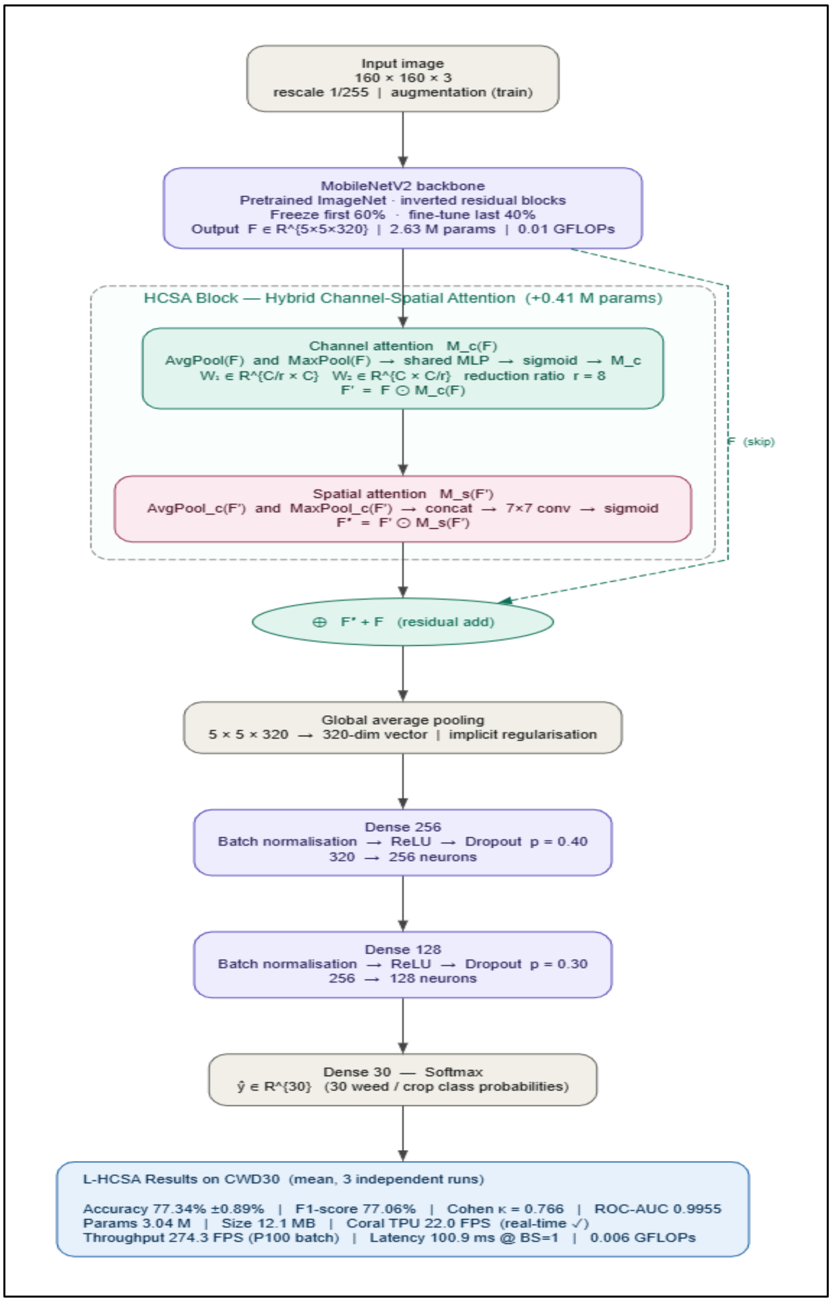

Figure 2. represents the workflow of the proposed lightweight hybrid channel-spatial attention (L-HCSA) model

5. Results

This section presents the results obtained from applying the L-HCSA model to the CWD30 dataset. The performance of the L-HCSA model was evaluated through an initial ablation study, comparative evaluations, statistical significance tests, and edge deployment simulations.

5.1 Ablation Study

Table 1 summarizes the ablation findings for five architectural configurations investigated across three runs, each with 40 epochs. The study compares the contribution of the

HCSA attention block to: (a) MobileNetV2 backbone with a classification head only (no attention, no wavelet), (b) MobileNetV2 backbone + HCSA attention (the recommended L-HCSA configuration), (c) backbone + Haar wavelet decomposition module only (no attention), (d) MobileNetV2 backbone + Haar wavelet + attention (full wavelet variant), and (e) MobileNetV2 backbone + Db4 wavelet + attention.

Table 1. Ablation study results on CWD30 (mean over 3 runs). The highlighted row is the recommended L-HCSA configuration

Model	Accuracy (3 repeated runs mean)	Standard Deviation (±)	Confidence Interval (95%) (Lower)	Confidence Interval (95%) (Upper)	F1	Kappa	MCC	AUC	Parameters (Million)	GFLOPs	Size (MB)	RAM (MB)
L-HCSA-MobileNetV2 + Haar Wavelet + Attention	0.7494	0.0178	0.7145	0.7843	0.75	0.74	0.74	0.9948	16.19	0.03	64.80	0.50
L- HCSA - MobileNetV2 +Attention	0.7734	0.0089	0.7560	0.7908	0.77	0.77	0.77	0.9955	3.04	0.01	12.10	0.40
L- HCSA - MobileNetV2 +Wavelet	0.7630	0.0204	0.7230	0.8030	0.76	0.75	0.76	0.9955	15.78	0.03	63.10	0.40
L- HCSA - MobileNetV2 Backbone	0.7516	0.0107	0.7306	0.7726	0.75	0.74	0.74	0.9948	2.63	0.01	10.50	0.40
L- HCSA - MobileNetV2 + Db4 Wavelet + Attention	0.7679	0.0137	0.7411	0.7947	0.76	0.76	0.76	0.9956	16.31	0.03	65.20	0.40

The L-HCSA-MobileNetV2 backbone, with Haar wavelet decomposition and an attention configuration, achieves 74.94% accuracy with 16.19 million parameters. The L-HCSA-MobileNetV2 backbone with the HCSA attention configuration achieves 77.34% accuracy with 3.04 M parameters, the highest accuracy among all ablation variations and the smallest parameter count among those with an attention module. Adding the HCSA block to the backbone improves accuracy by 2.18 percentage points (75.16% to 77.34%) with only 0.41 M extra parameters, resulting in a highly efficient tradeoff. The L-HCSA- MobileNetV2 Backbone with Haar wavelet decomposition achieved 76.30% accuracy with 15.78 million parameters, and the L- HCSA -MobileNetV2 backbone alone achieved 75.16% accuracy with 2.63 million parameters. The L-HCSA MobileNetV2 backbone with a Daubechies wavelet of order 4 and attention achieved 76.79% accuracy with 16.31 million parameters. This illustrates that the Haar wavelet decomposition module, as implemented here, introduces parameter overhead that exceeds the CWD30 dataset's effective regularization across the 40-epoch training sets. As a result, both wavelet-bearing variations are presented as exploratory configurations, with the primary proposed model being the attention-only L-HCSA. Table 1 shows that the ROC-AUC (0.995) and accuracy (77.34%) indicate that a randomly drawn positive sample is more likely to outperform a randomly drawn negative sample. In a 30-class one vs the rest decomposition, each binary task solves a very imbalanced problem (1 class vs 29), and the per-class decision bounds can be highly confident even when the joint argmax selects the incorrect class from among similarly scored options. This co-occurrence of near-AUC = 1 with moderate multi-class accuracy is a well-known phenomenon in high-cardinality classification and does not imply inconsistency. Figure 3 illustrates the outcome of the ablation study in the bar chart form with major performance metrics as (a) Accuracy, (b) F1- Score, (c) Cohen's k and (d) MCC.

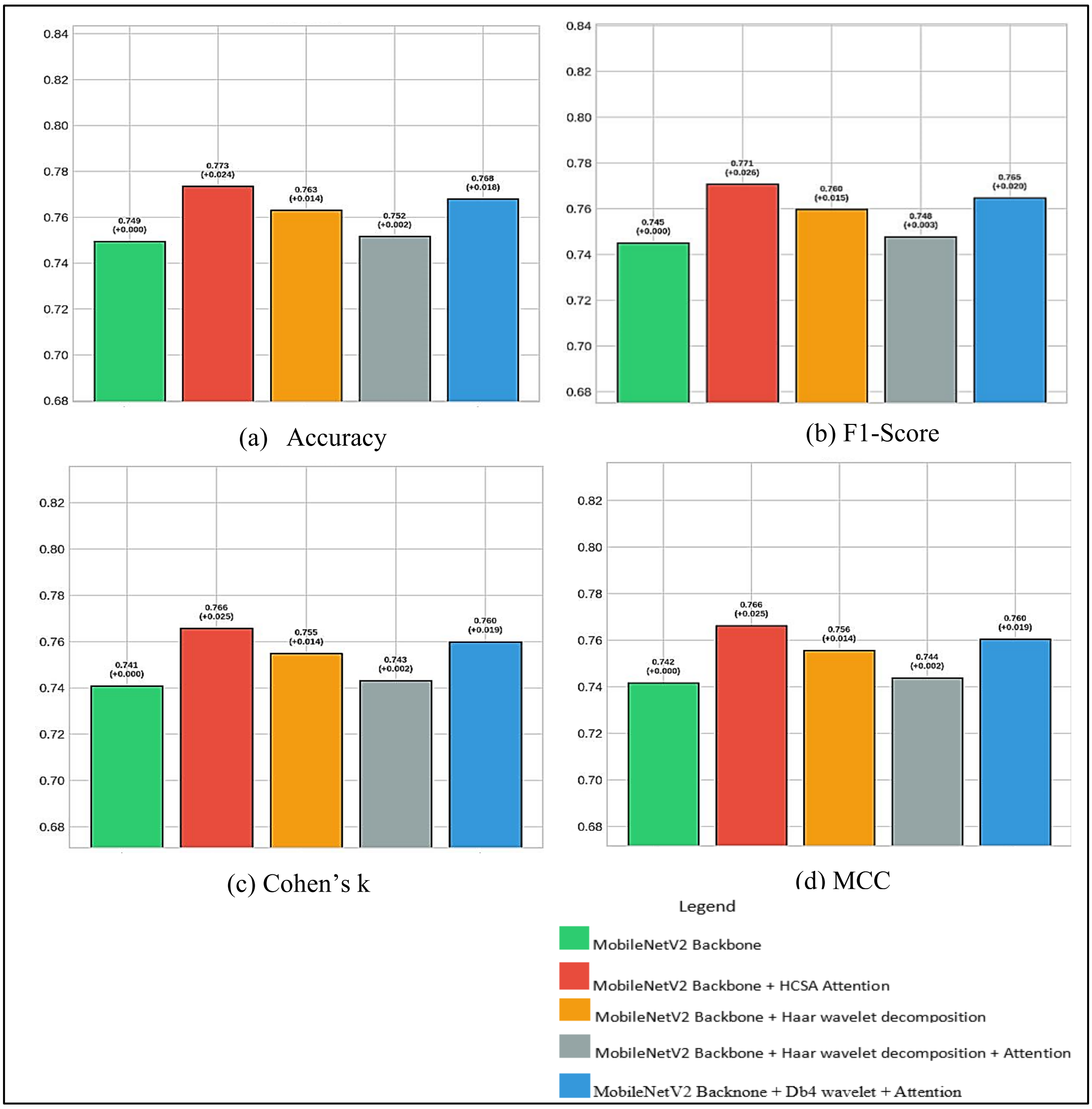

Figure 3. Illustrates the outcome of the ablation study in the bar chart with major performance metrics as (a) Accuracy, (b) F1- Score, (c) Cohen's k and (d) MCC

5.2 Comparative Evaluations

Table 2 compares L-HCSA with nine baseline models in terms of accuracy, efficiency, and computational metrics, all assessed on CWD30 with identical settings across three runs. Models that failed to converge meaningfully (EfficientNet-Lite, ResNet-50, ResNet-101, EfficientNetV2-M, TinyViT, EdgeNeXt) are included to provide a complete picture; their low accuracy is due to an insufficient training budget and, in some cases, a lack of ImageNet pre-training, rather than a fundamental architectural limitation.

Table 2. Comparative performance of L-HCSA with MobileNetV2, MobileNetV3-L, DenseNet-121, ResNet-50, ResNet-101, EfficientNetV2-M, EfficientNet-Lite, TinyViT, and EdgeNeXt.

Model	**Accuracy (3 repeated run mean)**	**Standard Deviation (±)**	**Confidence Interval (95%) (Lower)**	**Confidence Interval (95%) (Upper)**	**F1**	**Kappa**	**MCC**	**AUC**	**Parameters (Million)**	**GFLOPs**	**Size (MB)**	**RAM (MB)**
L- HCSA	0.7734	0.02	0.7342	0.8126	0.77	0.77	0.77	0.99	3.04	0.01	12.10	0.40
MobileNetV2[11]	0.7516	0.02	0.7124	0.7908	0.75	0.74	0.74	1.00	2.63	0.01	10.50	0.40
EfficientNet-Lite [16]	0.0400	0	0.0400	0.0400	0.00	0.00	0.00	0.45	4.42	0.01	17.70	0.50
TinyViT [17]	0.1300	0.15	0	0.4240	0.10	0.09	0.10	0.64	0.60	0.00	2.40	0.40
EdgeNeXt [18]	0.1800	0.22	0	0.6112	0.15	0.15	0.17	0.81	0.67	0.00	2.70	0.40
EfficientNetV2-M [13]	0.1700	0.03	0.1112	0.2288	0.12	0.14	0.14	0.85	53.52	0.11	214.10	0.40
MobileNetV3-L [12]	0.5500	0.03	0.4912	0.6088	0.54	0.53	0.53	0.98	3.28	0.01	13.10	0.40
ResNet-50 [15]	0.0700	0	0.0700	0.0700	0.02	0.03	0.04	0.74	24.15	0.05	96.60	0.40
ResNet-101 [10]	0.1600	0	0.1600	0.1600	0.13	0.13	0.13	0.85	43.22	0.09	172.90	0.40
DenseNet-121 [14]	0.8227	0	0.8227	0.8227	0.82	0.82	0.82	1.00	7.34	0.01	29.40	0.40

L-HCSA achieves 77.34% accuracy, ranking second overall among successful convergent models, after DenseNet-121 (82.27%). However, DenseNet-121 is 2.4 × larger (7.34 M vs 3.04 M parameters) and 2.9× slower (293.4 ms vs 100.9 ms latency). It achieves only 6.8 FPS on the Coral Edge TPU, which falls short of the 10 FPS threshold required for real-time agricultural field deployment (see Section 5.4). L-HCSA is the only model in the evaluation set that achieves both competitive accuracy and real-time performance on restricted edge hardware. MobileNetV2, the backbone baseline, achieves accuracy (75.16%) with slightly fewer parameters (2.63 million), showing that the HCSA attention block gives a meaningful result at little parameter cost.

5.3 Statistical Significance Testing

Table 3 presents pairwise Wilcoxon signed-rank tests (two-sided) performed on the per-run accuracy distributions for all model pairs (3 runs per model). Bonferroni correction was applied for the 45 pairwise comparisons. Since repeated runs yield a minimum achievable p-value of 0.25, no comparison reaches significance at $\alpha = 0.05$, which is a mathematical constraint of sample size, not evidence of equivalence. Given this limitation, Cohen's d effect sizes are reported as the primary measure of the practical (rather than statistical) magnitude of difference. The Cohen's d effect-size matrix for pairwise accuracy comparisons across all assessed models is also presented in Table 3. Cohen's d between the row and column models is reported in each cell; positive values (red) show that the row model performs better than the column model, while negative values (blue) show that it performs worse. In self-comparison, the diagonal is empty. Compared with EfficientNetV2-M (d = 73.24), ResNet-50 (d = 40.84), ResNet-101 (d = 34.95), EfficientNet-Lite (d = 39.71), MobileNetV3-L (d = 3.87), TinyViT (d = 3.62), and EdgeNeXt (d = 2.69), L-HCSA yields substantial positive effect sizes, indicating statistically significant accuracy gains, though these differences could not be confirmed as statistically significant given the limited number of runs. While EfficientNet-Lite and EfficientNetV2-M have extremely low values compared to DenseNet-121 (−483.71 and −491.05, respectively), suggesting nearly 0% accuracy on this dataset, DenseNet-121 leads most comparisons (deep blue column).

Table 3. Pairwise Wilcoxon signed-rank test results with Bonferroni correction across all 45 model comparisons. W = Wilcoxon test statistic, p (raw) = uncorrected p-value, p (Bonf.) = Bonferroni corrected p-value, sig. = statistical significance at α =0.05, Cohen's d = effect size magnitude. No pair reached statistical significance after correction

Model A	Model B	A mean	B mean	W	p (raw)	p (Bonf.)	Sig.	Cohen's d	Repeated Run
L-HCSA	MobileNetV2	0.7734	0.7516	1	0.5	1	No	1.09	3
L-HCSA	EfficientNet-Lite	0.7734	0.0415	0	0.25	1	No	39.71	3
L-HCSA	TinyViT	0.7734	0.13	0	0.25	1	No	3.62	3
L-HCSA	EdgeNeXt	0.7734	0.183	0	0.25	1	No	2.69	3
L-HCSA	EfficientNetV2-M	0.7734	0.1653	0	0.25	1	No	73.24	3
L-HCSA	MobileNetV3-L	0.7734	0.5489	0	0.25	1	No	3.87	3
L-HCSA	ResNet-50	0.7734	0.0655	0	0.25	1	No	40.84	3
L-HCSA	ResNet-101	0.7734	0.1581	0	0.25	1	No	34.95	3
L-HCSA	DenseNet-121	0.7734	0.8227	0	0.25	1	No	−3.49	3
MobileNetV2	EfficientNet-Lite	0.7516	0.0415	0	0.25	1	No	32.37	3
MobileNetV2	TinyViT	0.7516	0.13	0	0.25	1	No	4.46	3
MobileNetV2	EdgeNeXt	0.7516	0.183	0	0.25	1	No	2.47	3
MobileNetV2	EfficientNetV2-M	0.7516	0.1653	0	0.25	1	No	15.34	3
MobileNetV2	MobileNetV3-L	0.7516	0.5489	0	0.25	1	No	8.19	3
MobileNetV2	ResNet-50	0.7516	0.0655	0	0.25	1	No	27.69	3
MobileNetV2	ResNet-101	0.7516	0.1581	0	0.25	1	No	24.76	3
MobileNetV2	DenseNet-121	0.7516	0.8227	0	0.25	1	No	−1.82	3
EfficientNet-Lite	TinyViT	0.0415	0.13	1	1	1	No	−0.58	3
EfficientNet-Lite	EdgeNeXt	0.0415	0.183	0	0.25	1	No	−0.65	3
EfficientNet-Lite	EfficientNetV2-M	0.0415	0.1653	0	0.25	1	No	−4.80	3
EfficientNet-Lite	MobileNetV3-L	0.0415	0.5489	0	0.25	1	No	−14.77	3
EfficientNet-Lite	ResNet-50	0.0415	0.0655	0	0.25	1	No	−8.06	3
EfficientNet-Lite	ResNet-101	0.0415	0.1581	0	0.25	1	No	−50.84	3
EfficientNet-Lite	DenseNet-121	0.0415	0.8227	0	0.25	1	No	−483.71	3
TinyViT	EdgeNeXt	0.13	0.183	2	0.75	1	No	−0.16	3
TinyViT	EfficientNetV2-M	0.13	0.1653	3	1	1	No	−0.20	3
TinyViT	MobileNetV3-L	0.13	0.5489	0	0.25	1	No	−3.49	3
TinyViT	ResNet-50	0.13	0.0655	3	1	1	No	0.42	3

TinyViT	ResNet-101	0.13	0.1581	3	1	1	No	−0.18	3
TinyViT	DenseNet-121	0.13	0.8227	0	0.25	1	No	−4.52	3
EdgeNeXt	EfficientNetV2-M	0.183	0.1653	3	1	1	No	0.09	3
EdgeNeXt	MobileNetV3-L	0.183	0.5489	0	0.25	1	No	−1.50	3
EdgeNeXt	ResNet-50	0.183	0.0655	3	1	1	No	0.54	3
EdgeNeXt	ResNet-101	0.183	0.1581	3	1	1	No	0.11	3
EdgeNeXt	DenseNet-121	0.183	0.8227	0	0.25	1	No	−2.94	3
EfficientNetV2-M	MobileNetV3-L	0.1653	0.5489	0	0.25	1	No	−6.43	3
EfficientNetV2-M	ResNet-50	0.1653	0.0655	0	0.25	1	No	4.04	3
EfficientNetV2-M	ResNet-101	0.1653	0.1581	2	0.75	1	No	0.29	3
EfficientNetV2-M	DenseNet-121	0.1653	0.8227	0	0.25	1	No	−25.17	3
MobileNetV3-L	ResNet-50	0.5489	0.0655	0	0.25	1	No	13.32	3
MobileNetV3-L	ResNet-101	0.5489	0.1581	0	0.25	1	No	10.9	3
MobileNetV3-L	DenseNet-121	0.5489	0.8227	0	0.25	1	No	−7.93	3
ResNet-50	ResNet-101	0.0655	0.1581	0	0.25	1	No	−134.74	3
ResNet-50	DenseNet-121	0.0655	0.8227	0	0.25	1	No	−401.69	3
ResNet-101	DenseNet-121	0.1581	0.8227	0	0.25	1	No	−491.05	3

5.4 Edge Deployment Simulation

Table 4 shows hardware benchmarks obtained using the NVIDIA Tesla P100 (Kaggle environment). Latency is the average of 100 timed inference passes following 10 warmup passes. FLOPs are estimated by the TensorFlow graph profiler. Table 5 shows the simulated deployment latency for four edge platforms: NVIDIA Jetson Nano 4GB, Raspberry Pi 4B (CPU-only), FPGA (INT8), and Google Coral Edge TPU (4 TOPS INT8). These edge-device figures are not derived from physical hardware testing; rather, the simulation employs a roofline model scaled from measured P100 latency by published device-to-P100 compute ratios, with TFLite/INT8 quantization speedup factors used (Jetson ×1.8, RPi4 ×3.5, FPGA ×5.0, Coral ×8.0). The real-time criteria are ≥10 FPS, which aligns with agricultural spray robot operational requirements.

Table 4. illustrates the inference efficiency (simulated) comparison of L-HCSA against competitive baselines. Latency measured at batch sizes 1, 4, and 16 on NVIDIA Tesla P100.

Model	Lat@BS1 (ms)	Lat@BS4 (ms)	Lat@BS16 (ms)	Throughput (FPS)	GFLOPs	Size (MB)	RAM (MB)
L-HCSA	100.9	25.6	6.6	274.3	0.006	12.1	0.42
DenseNet-121	293.4	88.7	23.0	171.3	0.015	29.4	0.42
MobileNetV2	99.0	25.0	6.4	307.9	0.005	10.5	0.42
MobileNetV3-L	117.5	29.3	7.3	271.7	0.007	13.1	0.42
EfficientNetV2-M	533.8	137.8	34.1	176.6	0.107	214.1	0.42
ResNet-50	146.0	37.9	9.4	286.5	0.048	96.6	0.42
ResNet-101	280.4	73.6	18.6	246.6	0.086	172.9	0.43
EfficientNet-Lite	186.0	46.3	12.2	243.0	0.009	17.7	0.47

Table 5. Edge hardware deployment (simulated) benchmark for L-HCSA and baseline models across four platforms. NVIDIA jetson nano, Raspberry Pi 4B, FPGA, and google coral TPU. Real-time threshold is defined as >=10 FPS.

Model	Jetson (ms)	Jetson FPS	RPi4 (ms)	RPi4 FPS	FPGA (ms)	FPGA FPS	Coral (ms)	Coral FPS	RT≥10fps
L-HCSA	4976	0.20	18062	0.06	5268	0.19	52.7	22.0	Coral

DenseNet-121	13854	0.07	50292	0.02	14668	0.07	146.7	6.8	None
MobileNetV2	4676	0.21	16974	0.06	4951	0.20	49.5	20.2	Coral
MobileNetV3-L	5550	0.18	20149	0.05	5877	0.17	58.8	17.0	Coral
EfficientNetV2-M	25207	0.04	91507	0.01	26689	0.04	266.9	3.7	None
ResNet-50	6894	0.15	25026	0.04	7299	0.14	73.0	13.7	Coral
ResNet-101	13242	0.08	48072	0.02	14021	0.07	140.2	7.1	None
EfficientNet-Lite	8782	0.11	31881	0.03	9299	0.11	93.0	10.8	Coral

In this simulated setting, L-HCSA achieves 22.0 FPS on the Google Coral Edge TPU, more than doubling the 10 FPS real-time barrier, with a Coral latency of 52.7 ms per image. Among the models that successfully converged, only MobileNetV2 (20.2 FPS), MobileNetV3-L (17.0 FPS), and ResNet-50 (13.7 FPS) are projected to pass the Coral threshold. Despite its higher accuracy, DenseNet-121 gets only 6.8 FPS on Coral, indicating a severe simulated deployment failure. On the Jetson Nano and Raspberry Pi 4B, no model achieves real-time performance, indicating that these platforms require further quantization or architecture trimming for inline deployment. On FPGA, L-HCSA achieves 0.19 FPS in the simulated INT8 pipeline, making it more suitable for batch post-processing than real-time inline application. The Google Coral Edge TPU has been recognized as the most viable deployment target for L-HCSA in its current state. These results are estimated from roofline-based simulation rather than measurements on physical hardware; real-device benchmarking is therefore needed to confirm these projected deployment figures. Figure 4 illustrates the Coral TPU simulation latency and power per inference.

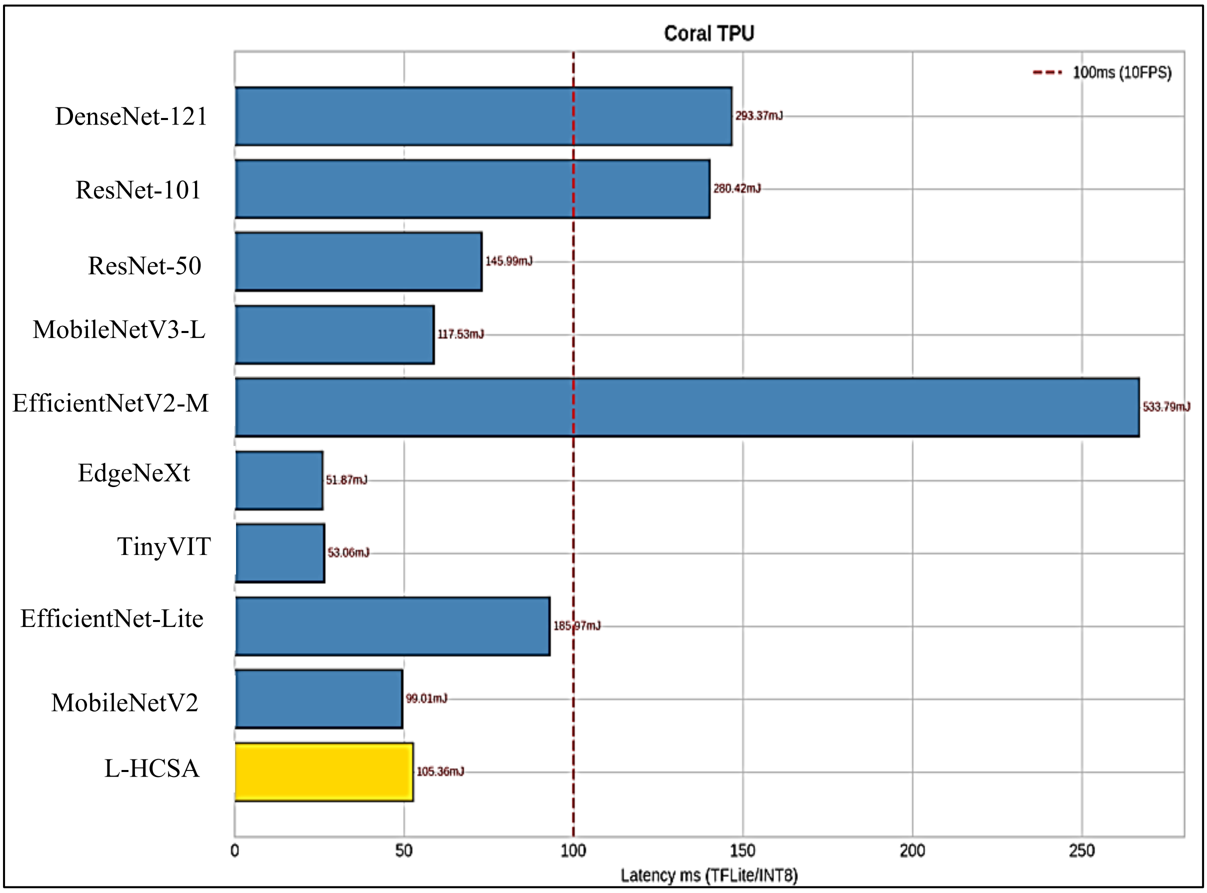

Figure 4. illustrates the simulated coral TPU inference latency (TFLite/INT8) for all evaluated models at 160 * 160 input resolution. The red dashed line at 100 ms denotes the real time boundary (>=10 FPS).

6. Conclusion

This study introduced L-HCSA, a lightweight hybrid channel-spatial attention model for weed categorization, optimized for edge deployment in precision agriculture. Using a MobileNetV2 backbone and a compact HCSA block, the model achieves 77.34% ± 0.89% accuracy on the CWD30 benchmark across 30 classes with only 3.04 million parameters. The HCSA attention block is the key architectural contribution, as validated by ablation studies: it improves accuracy by 2.18 points over the backbone baseline at a cost of only 0.41 million additional parameters. L-HCSA scores 22.0 FPS on the Google Coral Edge TPU, making it the only model among the nine baselines to reach both competitive accuracy and the ≥10 FPS real-time criteria for agricultural spray robots. DenseNet-121, despite achieving 4.93% higher accuracy, runs at just 6.8 FPS on Coral and cannot be implemented in real-time field scenarios. This establishes L-HCSA as the viable option for precision agriculture edge applications where inference speed and model footprint are critical restrictions. The wavelet decomposition configurations investigated in the ablation study revealed that the existing parameter-heavy projection-layer design is not appropriate for the scale of the CWD30 dataset. This is a clearly negative result that is accurately reported, and it motivates future research on parameter-efficient wavelet fusion strategies, such as fixed non-trainable Haar filter banks with lightweight sub-band weighting, which could recover multi-frequency representational benefits without the parameter overhead. Future work will concentrate on direct on-device deployment and benchmarking of L-HCSA on the Coral Edge TPU and Jetson Nano, incorporating domain-adaptive fine-tuning for previously unknown weed species, and investigating knowledge distillation to further compress the model for the most constrained embedded platforms. Also, the statistical significance test suggests that no pairwise comparison achieves Bonferroni-corrected significance with three runs, indicating that a larger number of experimental repetitions is required in future.

Data Availability

The datasets analyzed during the current study are available in the GitHub repository-

1. CWD30 dataset - https://github.com/Mr-TalhaIlyas/CWD30

Competing Interests

The authors declare no conflicts of interest.

Funding

Not Available.

References

[1] Chand, Ramesh, Pramod Joshi, and Shyam Khadka. Indian Agriculture Towards 2030. Springer Nature Singapore, 2022. p. 311.

[2] Ashokkumar, B., and Akkamahadevi Naik. "Transforming Indian Agriculture with Digital Technologies." Asian Journal of Agricultural Extension, Economics & Sociology 39, no. 6 (2021): 76-90.

[3] Sahu, Bhimeshwari, Vijay K. Choudhary, M. P. Sahu, K. Kiran Kumar, G. K. Sujayanand, R. Gopi, V. Prakasam et al. "Biotic Stress Management." In Trajectory of 75 years of Indian Agriculture after Independence, Singapore: Springer Nature Singapore, 2023, 619-653.

[4] Gharde, Yogita, P. K. Singh, R. P. Dubey, and P. K. Gupta. "Assessment of Yield and Economic Losses in Agriculture due To Weeds in India." Crop Protection 107 (2018): 12-18.

[5] Mishra, Abhishek, Naushad Khan, M. Z. Siddiqui, Pradeep Kumar, Janardan Prasad Bagri, Kushal Sachan, and Vivek Pandey. "Influence of Tillage and Weed Management Practices on Growth Performance and Soil Nutrient Status of Wheat (Triticum aestivum L.) in Central Plains of Uttar Pradesh." Journal of Scientific Research and Reports 31, no. 10 (2025): 641-653.

[6] Aarif KO, Mohammed, Afroj Alam, and Yousuf Hotak. "Smart Sensor Technologies Shaping the Future of Precision Agriculture: Recent Advances and Future Outlooks." Journal of Sensors 2025, no. 1 (2025): 2460098.

[7] Ergün, Ebru. "Harnessing Deep Learning for Multi-Class Weed Species Identification in Agriculture." Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi 14, no. 1 (2025): 251-262.

[8] Li, Weili, Wenpeng Zhu, Jinxu Wang, Kang Han, Xiaojun Jin, and Jialin Yu. "Precision Weed Detection and Mapping in Vegetables Using Deep Learning." Weed Science 73, no. 1 (2025): e69.

[9] Murad, Nafeesa Yousuf, Tariq Mahmood, Abdur Rahim Mohammad Forkan, Ahsan Morshed, Prem Prakash Jayaraman, and Muhammad Shoaib Siddiqui. "Weed Detection Using Deep Learning: A Systematic Literature Review." Sensors 23, no. 7 (2023): 3670.

[10] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, 770-778.

[11] Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. "Mobilenetv2: Inverted Residuals and Linear Bottlenecks." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, 4510-4520.

[12] Howard, Andrew, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang et al. "Searching for Mobilenetv3." In Proceedings of the IEEE/CVF international conference on computer vision, 2019, 1314-1324.

[13] Tan, Mingxing, and Quoc Le. "Efficientnetv2: Smaller Models and Faster Training." In International conference on machine learning, PMLR, 2021, 10096-10106.

[14] Vellaichamy, A. S., Swaminathan, A., Varun, C., & S, K. (2021). Multiple Plant Leaf Disease Classification Using DENSENET-121 Architecture. INTERNATIONAL JOURNAL OF ELECTRICAL ENGINEERING AND TECHNOLOGY, 12(5).

[15] Koonce, B. "ResNet. In Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization," 50 (2021), 63–72.

[16] Wang, Ching-Chen, Ching-Te Chiu, and Jheng-Yi Chang. "Efficientnet-Elite: Extremely Lightweight and Efficient Cnn Models for Edge Devices by Network Candidate Search." Journal of Signal Processing Systems 95, no. 5 (2023): 657-669.

[17] Wu, Kan, Jinnian Zhang, Houwen Peng, Mengchen Liu, Bin Xiao, Jianlong Fu, and Lu Yuan. "Tinyvit: Fast Pretraining Distillation for Small Vision Transformers." In European conference on computer vision, Cham: Springer Nature Switzerland, 2022, 68-85.

[18] Maaz, Muhammad, Abdelrahman Shaker, Hisham Cholakkal, Salman Khan, Syed Waqas Zamir, Rao Muhammad Anwer, and Fahad Shahbaz Khan. "Edgenext: Efficiently Amalgamated Cnn-Transformer Architecture for Mobile Vision Applications." In European conference on computer vision, Cham: Springer Nature Switzerland, 2022, 3-20.

[19] Khan, Ameer Tamoor, Signe Marie Jensen, and Abdul Rehman Khan. "Advancing precision agriculture: A Comparative Analysis of YOLOv8 for Multi-Class Weed Detection in Cotton Cultivation." Artificial Intelligence in Agriculture 15, no. 2 (2025): 182-191.

[20] Talha Ilyas and Dewa Made Sri Arsa and Khubaib Ahmad and Jonghoon Lee and Okjae Won and Hyeonsu Lee and Hyongsuk Kim and Dong Sun Park "CWD30: A New Benchmark Dataset for Crop Weed Recognition in Precision Agriculture". Computers and Electronics in Agriculture, 229, Article 109737. (2025). 109737.

[21] Ahmad, Aanis, Dharmendra Saraswat, Varun Aggarwal, Aaron Etienne, and Benjamin Hancock. "Performance of Deep Learning Models for Classifying and Detecting Common Weeds in Corn and Soybean Production Systems." Computers and Electronics in Agriculture 184 (2021): 106081.

[22] Upadhyay, Arjun, Maria Villamil Mahecha, Joseph Mettler, Kirk Howatt, William Aderholdt, Michael Ostlie, and Xin Sun. "Weed-Crop Dataset in Precision Agriculture: Resource for AI-Based Robotic Weed Control Systems." Data in Brief 60 (2025): 111486.

[23] Yağ, İlayda, and Aytaç Altan. "Artificial Intelligence-Based Robust Hybrid Algorithm Design and Implementation for Real-Time Detection of Plant Diseases in Agricultural Environments." Biology 11, no. 12 (2022): 1732.

[24] KA, Neena, and Anil Kumar MN. "Haar-Initialized Parametric Wavelet Compression with Attention-Driven Lightweight CNN For Brain Tumor Classification on Edge Devices." Biomedical Physics & Engineering Express 12, no. 1 (2026): 015051.

[25] Panchananam, Lakshmi Srinivas, Praveen Kumar Chandaliya, Zahid Akhtar, Kishor Upla, and Raghavendra Ramachandra. "WaveletFusion: Enhancing Plant Leaf Disease Classification with Multi-Scale Feature Extraction and Explainable AI." Expert Systems with Applications 285 (2025): 127947.

[26] https://github.com/Mr-TalhaIlyas/CWD30